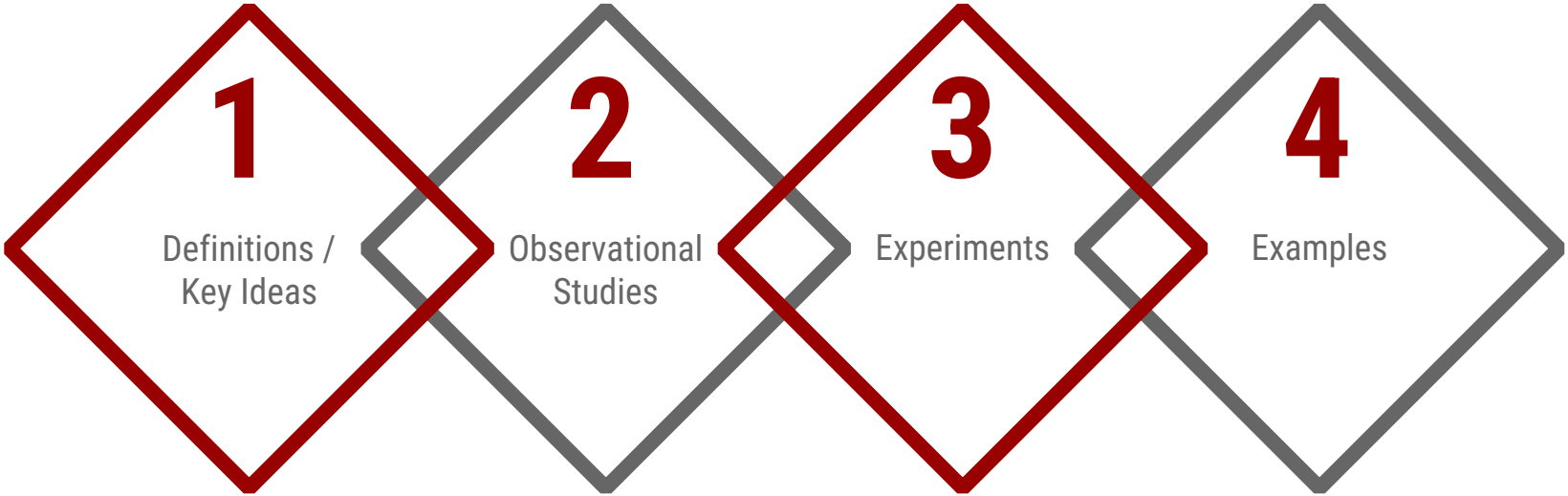
11.1 Statistical Studies

Goals for the Day



1

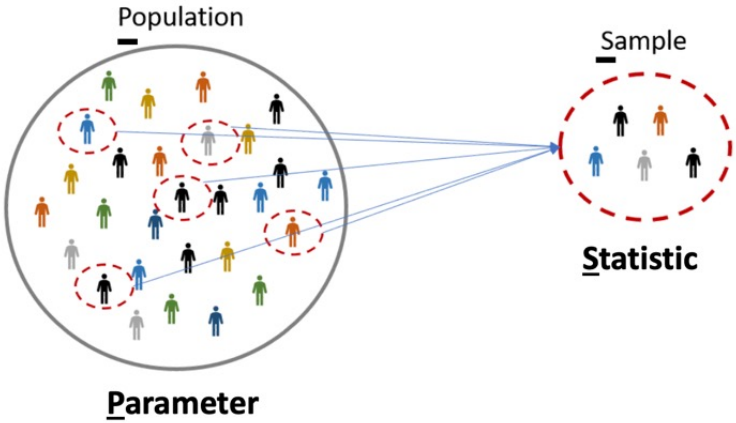
Definitions / Key Ideas

What is Statistics?

Statistics is the science of gathering, describing, and analyzing data.

Population vs Sample

- **Population** – the particular group of interest in a study
  - The set of all individuals/objects of interest
- **Sample** – a subset of the individuals/objects from the population of interest



Population vs Sample

Example

Let's say we want to know if Indiana is a cat or dog state.

What is the population? Every person in ALL of Indiana

What is the sample? Everybody in Muncie ONLY

Parameter vs Statistic

- (Population) Parameter – a fixed numerical value that describes the population
  - EX: Percentage of people in all of Indiana who prefer cats
  - Would need to take a **census** (ask everyone in the population) to know this value (or estimate it)
- (Sample) Statistic – a numerical value that describes the sample that can vary from sample to sample
  - EX: Percentage of people in Muncie who prefer cats (will be different than for Indy)

Observational Study vs Experiment

- An **observational study** observes existing data.
  - Does not impose any restraints (no random assignment).
  - Can reveal association or correlation between variables, but not causation.
- An **experiment** generates data to help identify cause-and-effect relationships.
  - Imposes treatments and controls randomly to groups.
  - More accurate to determine a relationship between the explanatory variable and response.

2

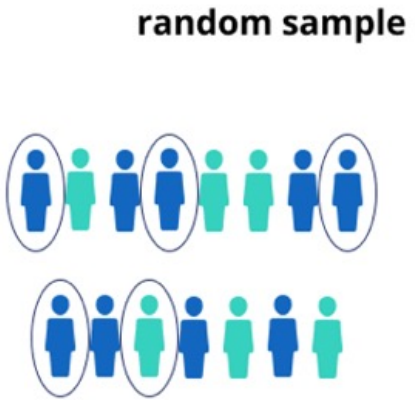
Observational Studies

Sampling

- In the context of observational studies, we need to find a way to pick who/what is going to be included in the sample.
- Goal: A **representative sample** – a sample that has the same relevant characteristics as the population AND does not favor one group of the population over another.
- Many different methods for sampling.

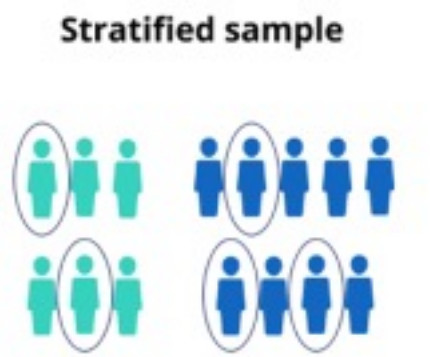
Sampling Techniques

- **Random Sample** – every member of the population has an equal chance of being selected
  - This is generally desirable but can be difficult to achieve.



Sampling Techniques

- Stratified Random Sample – dividing the population into homogeneous (similar characteristics) groups
  1. Stratify the population - divide the population into similar groups (e.g., based on age or gender)
  2. Take random sample from each group (strata)
  3. Combine the groups from each strata to form your sample

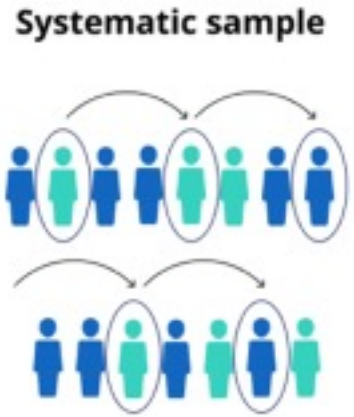


Sampling Techniques

- Cluster Sample – dividing the population into mini-populations. This gives us an unbiased sample and is often a more practical / affordable method.
  1. Split the population into representative groups called clusters (should resemble overall population)
  2. Use random sampling to select several whole clusters
  3. Perform a census of each selected (collect data from every member).

Sampling Techniques

- **Systematic Sample** – selecting every n^{th} member of the population



Sampling Techniques

- Convenience Sample – include individuals who are convenient to sample (for the researcher); AVOID!
  - This group may not be representative of population
  - Frequently leads to biased results

Example #1

Describe how you could obtain a sample to answer the question below using each of the following types of sampling methods.

I want to determine the proportion of MATH 125 students that has a Mac laptop.

Random – Randomly ask 10 students from class

Cluster – Randomly sample 3 tables; ask everyone in that (take a census of the) table

Convenience – Ask the 5 students closest to me

Stratified – Randomly sample 5 students from the list of Freshman and Sophomores respectively

Systematic – choose every 4^{th} student off the roster

3

Experiments

Definitions

- Explanatory Variable
  - The variable(s) being used to **explain** a response (the variable doing the explaining).
- Response Variable
  - The variable we are ultimately interested in (the one trying to be explained).
- *Our GOAL is to see if we can describe our response with an explanatory variable(s).*

More Definitions

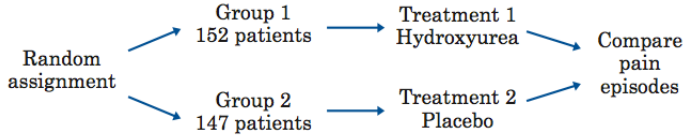
- **Subjects** (also called **Experimental Units**) are what the *treatment* is being *applied* to.
- **Treatment** is the *experimental condition.*

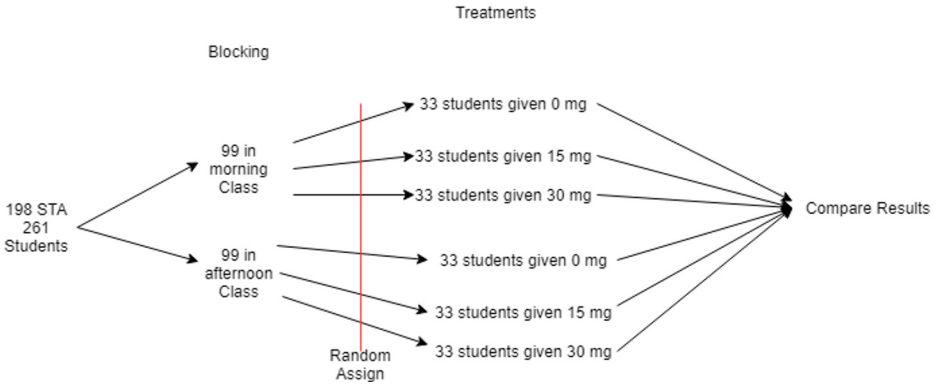
Different Groups in an Experiment

- **Treatment Groups –** The group the receives the treatment.
- Control Group
  - The group that does not receive the treatment.
  - We have a control group to compare our treatment groups to.
  - We have to know what the “base” level of our response is.

Principles of Experimental Design

1. Randomize the control and treatment groups.
   - Equalize effects of known and unknown variation.
2. Control for outside effects on the variable.
   - Make conditions as similar as possible for all groups so that the only difference is the imposed treatment
3. Replicate the experiment a significant number of times to see meaningful patterns.
   - Apply each treatment to a number of subjects.





More Terms

- Experiments can get complicated, lots of things to consider. These include but are not limited to:
  - Placebo and its effects
  - Single- or double-blind studies
  - Confounding variables
  - Blocking

Example #2

Describe how you could obtain a sample to answer the question below using each of the following types of sampling methods.

You are tasked with conducting a survey to answer the question, “What is the favorite subjects of students who attend East High School?”

Random – Get a list of all student names and randomly select 10 names

(could number each name and randomly generate 10 numbers)

Stratified –Divide students by grade level OR by social group band, football, etc.) and then randomly sample within each group

(Students are the same WITHIN each group, but different ACROSS groups)

Cluster – Randomly sample 5 classrooms OR buses, and then ask every student in each.

(Should be a mix of students in each cluster, all clusters same)

Systematic – Ask every 5th student that arrives in the morning.

Get a list of all students names and select every 10th name.

Convenience – Walk through hallway and ask first 20 students you pass.

(Bad because maybe I am by the chem lab and only ask chem students; their opinions might not match the overall student body’s opinion

Example #3

Identify the response variable, subjects, and treatment for the following experiment:

The MATH 125 team is interested if the amount of sleep affects students' memory of a 5-minute video shown in class. 25 students were randomly chosen to sleep for 6 hours, another 30 were randomly chosen to sleep for 8 hours; the rest of the class (28 students) were assigned to sleep for 10 hours. Students were then scored on the amount they were able to remember on the video.

Response variable – Score on video

Subjects – Students

Treatment – Amount of sleep